The speakers' accent shapes the listeners' phonological predictions during speech perception

Angèle Brunellière [a],[*], Salvador Soto-Faraco [b]

[a] *Unité de Recherche en Sciences Cognitives et Affectives, University of Lille 3, France*
[b] *ICREA & Departament de Tecnologies de la Informació i les Comunicacions, Universitat Pompeu Fabra, Spain*

A B S T R A C T

This study investigates the specificity of predictive coding in spoken word comprehension using event-related potentials (ERPs). We measured word-evoked ERPs in Catalan speakers listening to semantically constraining sentences produced in their native regional accent (Experiment 1) or in a non-native accent (Experiment 2). Semantically anomalous words produced long-lasting negative shift (N400) starting as early as 250 ms, thus reflecting phonological as well as semantic mismatch. Semantically expected but phonologically unexpected (non-native forms embedded in a native context) produced only an early (∼250 ms) negative difference. In contrast, this phonological expectancy effect failed for native albeit phonologically unexpected target words embedded in a non-native context. These results suggest phonologically precise expectations when operating over native input, whilst phonologically less specified expectations in a non-native context. Our findings shed light on contextual influence during word recognition, suggesting that word form prediction based on context is sensitive and adaptive to phonological variability.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

In everyday conversations, words are embedded in the context of a message with a particular phonological, semantic, syntactic and paralinguistic context. Mounting evidence suggests that this context can strongly constrain the interpretation of incoming speech input and hence influence the ease with which words are recognized (e.g., DeLong, Urbach, & Kutas, 2005; Frauenfelder & Tyler, 1987; Pickering & Garrod, 2007; Tanenhaus & Lucas, 1987). But can such constraining mechanisms operate at a phonologically detailed level? That is, do the constraints adapt to phonological variations in the speakers' characteristics, such as for example, a regional accent? Or are they, instead, tied to more abstract but less constraining levels of representation? This is precisely the question addressed here.

Interactive models of word recognition propose a pre-activation[1] at a lexical level, whereby likely word candidates according to the context constrain input analysis at early stages of word recognition (Marslen-Wilson & Welsh, 1978; McClelland & Elman, 1986;

Morton, 1979). Recent studies using measures with high temporal resolution (such as electrophysiology and eye movements) provide experimental evidence for the role of on-line expectations in real-time language processing (Altmann & Kamide, 1999; Delong et al., 2005; Kamide, Altmann, & Haywood, 2003; van Berkum, van den Brink, Tesink, Kos, & Hagoort, 2008; Wicha, Bates, Moreno, & Kutas, 2003a, 2003b; Wicha, Moreno, & Kutas, 2004). For example, eye tracking studies show that listeners anticipate upcoming words on the basis of semantic properties of verbs and preverbal arguments (Altmann & Kamide, 1999; Kamide et al., 2003). Event-related potentials (ERPs), which offer a real-time measure of the dynamics of neuronal responses, have also supported a role of expectancies for upcoming word forms driven by the sentential context (Delong et al., 2005; van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005; Wicha et al., 2003a, 2003b, 2004). For example, ERP studies manipulating gender agreement rules in semantically constraining sentences have demonstrated strong prediction effects for upcoming words in sentential context, as well as for the syntactic properties of expected words (in Spanish between articles and nouns, Wicha et al., 2003a, 2003b, 2004, in Flemish between adjectives and nouns, van Berkum et al., 2005).

An important question about these mechanisms concerns the specificity of the predicted word forms during on-line sentence processing. Given the top-down nature of expectations for specific word forms, this expectation can potentially constrain input analysis at very early stages of word recognition. Delong et al. (2005)

---

* Corresponding author. Address: Unité de Recherche en Sciences Cognitives et Affectives, Université Charles-de-Gaulle Lille 3, Domaine Universitaire du Pont de Bois, BP 149, 59653 Villeneuve d'Ascq Cedex, France.
E-mail address: angele.brunelliere@univ-lille3.fr (A. Brunellière).

[1] In interactive models, lexical candidates are said to be pre-activated such that the input analysis is constrained without the involvement of checking processes (i.e. the input checked against the prediction).

attempted to explore this question in written sentences by investigating the N400 response, an electrophysiological component whose amplitude reflects lexical access. In particular, they measured the amplitude of the N400 wave on the presentation of the English indefinite article "a" or "an." These two forms of the article are identical in their semantic and syntactic properties, differing only in phonological form, which depends on the initial sound of the following noun (e.g. "a kite", "an airplane"). The authors found that the amplitude of the N400 elicited by the article decreased when the predictability of the expected, but not yet presented, noun increased. The observation of this expectancy effect from the article thus suggests that the context can lead to the generation of top-down expectations for specific word forms.

It is clear, then, that the degree of word form specificity at which context can produce expectations goes down to at least the phonological level. However, given the problem of the lack of invariance in the speech signal (see Liberman, Cooper, Harris, & MacNeilage, 1962; McClelland & Elman, 1986; Stevens & Halle, 1967), an intriguing question is whether these predictions operate at a more specific level of representation where a detailed form of the input is expected, or whether they instead remain confined to an abstract representation that neglects phonological variability of word realizations across, for example, different speakers of a language. The present study was designed to explore a common linguistic situation related to this issue by addressing whether listeners' expectations about regional accents reflect the fine detail of word forms on the basis of sentential context. In particular, we examined whether listeners' expectations include the particular phonological forms of a likely word in the regional accent of the speaker.

Answering this question will allow us to determine whether highly specified predictions of word forms constrain input analysis in a flexible, form-oriented manner (thus leading to more efficient word recognition). This question partly hinges upon the classic controversy of how words are represented in the mental lexicon (Goldinger, 1996, 1998; Marslen-Wilson, 1984; Morton, 1979; Norris, McQueen, & Cutler, 2003). Among the theories of spoken word recognition, Goldinger (1996, 1998) has proposed that word representations are stored with a high level of detail about sound shape, such as the talker's voice. According to this account, words are encoded in memory as exemplar-based representations of fine-grained phonological cues present in the auditory input. According to this view, the encoding of fine-grained phonological cues usually occurs at phonetic level across the variability in pronunciation between speakers. An interesting and curious phenomenon supporting the processing of fine acoustic detail is observed in conversational situations, wherein talkers unconsciously tend to mimic their interlocutors' pronunciation based on fine-grained information in the speech input. For instance, speakers tend to match their partner's vocal intensity (Natale, 1975), speech rate (Giles, Coupland, & Coupland, 1991) and the phonetic realizations of words (Pardo, 2006). This observation thus runs in favour of the storage of exemplar-based, detailed representations of the sound shapes of words. In stark contrast to this framework, some authors have argued that words are represented as abstract phonological codes (e.g. Marslen-Wilson, 1984; Morton, 1979; Norris et al., 2003). In this kind of abstract representations models, variability in the speech input (e.g., inter-speaker variability in voice, speech rate, dialect-dependent phonological and phonetic realizations) is treated as irrelevant variation and removed in the activation and selection of lexical candidates. According to this account, this is accomplished using a filter that can match the incoming signal to abstract representations in the lexicon, via a normalization process. Thus, to a certain extent, it may be inferred in relation to our objective that one main distinction between the two frameworks might reside in the question of whether or not multiple variants of a single word are stored. More precisely, this is, beyond, the question of whether the encoding encompasses phonological variability or else, it does so down to specify phonetic variability too.

1.1. Scope of the study

In the present study, we seek to gauge the specificity at which generation of expectancies for spoken word forms can be expressed in terms of specific phonological variants. To this end, we studied real phonological variation between regional accents and exploited the capacity of event-related potentials (ERPs) to study brain processes in real time with high temporal resolution. In particular, we focused on ERP components observed to be sensitive to semantic and word form constraints in sentential context. The well-established N400 is a negative wave that peaks around 400 ms after stimulus onset and typically has a posterior distribution across the scalp. In sentential contexts, the N400 reflects the consequences of contextually based expectations regarding upcoming words at a lexico-semantic level. The amplitude of the N400 is smaller for highly predictable words compared to contextually coherent but less predictable words in meaningful sentential context (e.g., Delong et al., 2005; Kutas & Hillyard, 1984). Words that are less predictable or semantically anomalous given the sentential context elicit a large N400 component. Another, earlier negative wave, called N200, has also been described as an index of contextual generation of expectations for upcoming words. But, in contrast to the N400, the N200 is thought to mostly reflect the consequences of expectancies at a phonological level (Connolly & Phillips, 1994; Connolly, Phillips, Stewart, & Brake, 1992; Connolly, Stewart, & Phillips, 1990). In a seminal study, Connolly et al. (1992) found a decrease of negativity, peaking around 200 ms, in addition to the decrease in N400 amplitude, after words presented in strongly constraining sentences relative to words in weakly constraining sentences. Connolly et al. (1992) suggested that the N200 is triggered by a mismatch between the initial phonemes of the incoming word and the initial phonemes of the words activated by expectations on the basis of the preceding context (see also Connolly & Phillips, 1994; Van den Brink, Brown, & Hagoort, 2001). Taken together, the results are consistent with the functional interpretation that the N200 reflects mismatches, occurring at a phonological level, between the incoming word and the expected word from the preceding context. Note, however, that some studies (van den Brink & Hagoort, 2004; Van Petten, Coulson, Rubin, Plante, & Parks, 1999) have described difficulties separating this early N200 negativity from classical N400 effects. In particular, Van Petten et al. (1999) showed that the onset of N400 effects is related to the moment at which the incoming signal suffices to register a mismatch.

To investigate whether listeners predict detailed phonological word forms of expected words according to context, we capitalized on the rule of vowel reduction in Catalan, which is applied in some but not all regional accents. Specifically, Eastern Catalan (spoken in Barcelona) applies vowel reduction (e.g., /a/ and /e/ segments become a schwa sound /ə/ in unstressed syllables), whereas Western Catalan (spoken in Lleida, for example) does not apply vowel reduction. That is, the rule of vowel reduction in unstressed syllables leads to vowel neutralizations in Eastern Catalan (Alarcos, 1953). For example, the word third in Catalan is pronounced /tersé/ in a Western Catalan accent, whereas it is produced /tərsé/ in Eastern Catalan. In a similar fashion, /o/ and /ɔ/ are reduced to /u/ in unstressed syllables, so that the word chocolates is pronounced /bombóns/ in Western Catalan, whereas it is produced /bumbóns/ in Eastern Catalan. These regional variations in vowel realization in Catalan are thus based on distinct phonemes, not simply on allophonic variation. Therefore, it is important to clarify that this manipulation does not directly address whether predic-

Table 1
Examples of experimental conditions in Experiments 1 and 2.

Conditions	Regional accent of context	Regional accent of final word	Examples
Experiment 1			
Fully expected	Native	Native	És una família molt estricta, abans d'aixecar-te de la taula has de demanar /pərmís/
			It is a very strict family, before getting up from the table, you have to ask the permission
Phonologically unexpected	Native	Non-native	És una família molt estricta, abans d'aixecar-te de la taula has de demanar /p**e**rmís/
			It is a very strict family, before getting up from the table, you have to ask the permission
Fully unexpected	Native	Native	*És una família molt estricta, abans d'aixecar-te de la taula has de demanar /p**ikáda**/
			*It is a very strict family, before getting up from the table, you have to ask the bite**
Experiment 2			
Fully expected	Non-native	Non-native	És una família molt estricta, abans d'aixecar-te de la taula has de demanar /permís/
			It is a very strict family, before getting up from the table, you have to ask the permission
Phonologically unexpected	Non-native	Native	És una família molt estricta, abans d'aixecar-te de la taula has de demanar /p**ə**rmís/
			It is a very strict family, before getting up from the table, you have to ask the permission
Fully unexpected	Non-native	Non-native	És una família molt estricta, abans d'aixecar-te de la taula has de demanar/p**ikáda**/*
			*It is a very strict family, before getting up from the table, you have to ask the bite**

[*] Semantic incongruence, phonological mismatches in red. *Native*: Eastern Catalan accent, Non-native: Western Catalan accent.

tions specify allophonic variation or not, corresponding to variability at phonetic level. Instead, the interpretation concerns the expression of phonological variation by predictive mechanisms.

We presented native Eastern Catalan speakers with semantically constraining sentences produced either in their native regional accent (Experiment 1) or in the alternative non-native regional accent, i.e. Western Catalan (Experiment 2). In each experiment, the words ending the sentence frame differed from the second phoneme onwards, creating three experimental conditions. In the fully expected condition, the final word was the most expected word provided the prior semantic context and had an expected phonological form according to the regional accent of the context. In the phonologically unexpected condition, the final word was the most expected word provided the prior context but had an unexpected, but legal, phonological form with respect to the regional accent of the context. In this condition, the expected phonological word form was replaced by the phonological variant produced in the alternative regional accent. In the fully unexpected condition, the word final was semantically incongruent to the prior context (and hence by definition its phonological form was also unexpected). Examples of experimental stimuli are presented in Table 1.

The critical condition was the phonologically unexpected one. If predicted word forms incorporate details of the regional accent produced in the framing context, then we should observe increased amplitude of early effects elicited by the phonological mismatch between the fully expected and phonologically unexpected conditions (putatively, the N200 or early N400 effects). Since these two conditions are semantically equivalent thereafter, we do not expect later N400 modulations in this comparison. Additionally, and in line with prior studies, the fully unexpected condition should produce early differences in comparison to the fully expected condition, as per the phoneme mismatch with the expected form, and also late effects as per the lexical and semantic mismatch with the expected word. All predictions were based on differences between the experimental conditions and interactions between the experimental conditions and the scalp distribution of the two ERP components of interest.

2. Experiment 1: Phonological expectations in native regional accent

The objective of the Experiment 1 was to examine whether the prediction of detailed phonological word forms arises in the context of the native regional accent of the listener.

2.1. Methods

2.1.1. Participants

Twenty-one dominant Catalan-speaking students from the Pompeu Fabra University, aged between 19 and 25 years (13 female, mean age: 21, SD age: 1.9) were selected. None had hearing or language impairments according to self report. Catalan was the only language spoken with their family and the most frequent language spoken in their everyday life. The participants came from the eastern part of Catalonia, mainly from Barcelona and the surrounding area, and their parents were native Eastern Catalan speakers as well. All were right-handed as assessed by the Edinburgh handedness inventory (Oldfield, 1971). They received monetary compensation for participation (8€/h). Before the beginning of the experiment, participants gave their written informed consent.

2.1.2. Materials

The experimental stimuli consisted of a set of 135 triplets of semantically constraining sentences. The sentences in each triplet were spoken in the Eastern Catalan (native) accent up until the penultimate word (context) and differed in the final word (target). In the fully expected condition, the final word was the most expected word provided by the prior context and its phonological form was in agreement with the accent in the context (i.e., native regional accent, in this case). In the phonologically unexpected condition, the final word was the most expected word provided by the prior context but its phonological form was not in agreement with the accent used in the context (i.e. non-native regional accent, Western Catalan). In the fully unexpected condition, the final word was semantically unexpected in the prior context. In all cases, the manipulation of the final word involved variation starting from the second phoneme (always a vowel as the nucleus of the first syllable).

The selection of the 135 triplets was based on the classical cloze procedure, in which 30 Catalan speakers (different from the ones tested in the ERP experiment) were asked to complete the sentence fragments with the first word that comes to mind. The 135 triplets had a cloze probability of at least 0.50 (mean: 0.78; range: 0.50–1). Because the sentences in each triplet differed only in the final word, the context fragments in each triplet were matched for length (mean: 13.7 words, range: 7–22) and for markers[2] of regional accent (mean: 6.4 phonemes, range: 3–15). In the phonologically

---

[2] Number of phonemes allowing detection of the regional accent that the context is delivered in.

unexpected condition, the final word differed from the expected word only in its second phoneme, which always indicated that the word was produced in the non-native Western accent (see Table 1). The phonological variation was produced by variation in two phonemic categories (/ə/ vs. /e/ or /u/ vs. /o/). In the fully unexpected condition, the final word differed from the expected word in the second phoneme, just as in the phonologically unexpected condition, but in this case the second phoneme was always /i/, which is not subject to regional phonological variations in Catalan. In this last condition, the remaining phonemes of the target word were additionally not shared with the expected word to produce a semantically anomalous word. The expected and unexpected words were selected from the Catalan Dictionary of Frequencies database (Rafel i Fontanals, 1998) and matched for frequency and number of phonemes (819 vs. 794 tokens per million and 5.8 vs. 6.0 phonemes long, respectively). The expected and unexpected words were also matched for lexical category (noun, verb, adverb, and adjective). We selected final words beginning with an unstressed syllable, on which regional variation in Catalan occurs (vowel reduction or not). The final target words were 2 or 3 syllables long and were not subject to regional phonological variations on the second or third syllable. All final words began with a plosive segment (/p/, /t/, /b/, /d/) to provide a clear physical marker on the spectrogram, which made it possible to precisely align the onsets of targets for the ERP recordings.

To avoid repeated presentation of the same sentence or final word, only one sentence in each triplet was presented to a given participant. Three equivalent experimental lists of 45 trials per condition were constructed so that each participant was exposed to all experimental conditions across sentences (fully expected, phonologically unexpected, fully unexpected). The different combinations of sentence conditions were counterbalanced across experimental lists. None of the final words was present in the sentential context of any sentence. In addition to the experimental stimuli, incongruent and congruent sentence filler trials were created to prevent the adoption of strategies based on the second phoneme due to semantic incongruence. These fillers included: 45 sentences ending with a semantically incongruent word wherein the violation was evident at the first phoneme; ninety sentences ending with a semantically expected word to balance the overall number of sentences that were semantically incongruent and congruent throughout the materials.

For recording, all stimuli were produced several times by a dominant Catalan-speaking female with standard Barcelona accent (i.e. Eastern Catalan accent) and were digitized at a sampling rate of 44 kHz with 16-bit. The speaker was asked to pronounce the sentences with natural prosody at normal speaking rate. To make sure that intonation and speaking rate were kept constant between experimental conditions, the experimental stimuli were recorded in triplets. The order of experimental stimuli within each triplet was counterbalanced to avoid an effect of first reading in a particular experimental condition. To create the phonologically unexpected condition, the speaker was asked to end the sentence by applying the phonological features of Western Catalan when the final word was produced. The selection of auditory sentences was based on natural intonation and speaking rate, and correct pronunciations of phonological features of Western Catalan on the final word, creating the phonologically unexpected condition. The total mean duration of sentences before the final word was 3773 ms (SD: 1000) for the fully expected condition, 3792 ms (SD: 954) for the phonologically unexpected condition and 3777 ms (SD: 957) for the fully unexpected condition. The mean duration of the final word was 463 ms (SD: 92) for the fully expected condition, 475 (SD: 94) for the phonologically unexpected condition and 478 (SD: 84) for the fully unexpected condition. The total duration of sentence context (up to the onset of the final word)
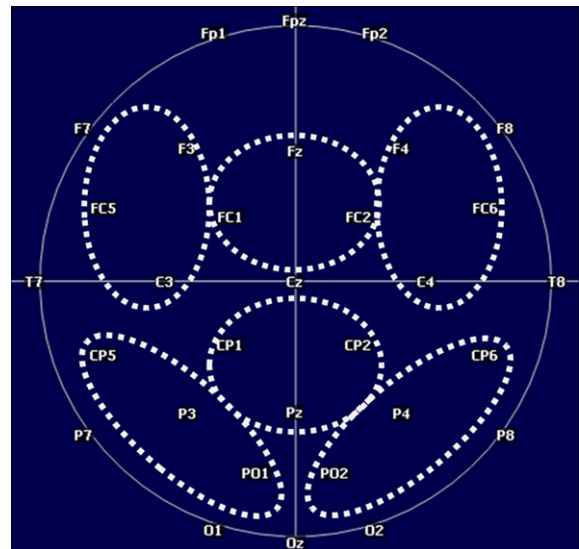

Fig. 1. Distribution of electrodes on the scalp. In a dotted line, the groupings of electrodes used as factors in the analysis of variance (see main text).

and that of the final word (target), was similar across the experimental conditions ($F(2, 268) = 0.43$, $p > .2$; $F(2, 268) = 1.81$, $p = 0.18$, respectively).

2.1.3. Experimental procedure

Each trial started with a fixation cross in the middle of the screen, which was presented 500 ms before the onset of the auditory sentences and remained there until 1000 ms afterwards. The sentences were presented binaurally at a constant, comfortable sound level pressure via headphones. The next trial began 2000 ms after the end of the previous trial. To minimize muscular artefacts, participants were asked to fixate the fixation cross and not move their eyes, and to make any necessary movements for comfort when the fixation cross disappeared. Participants first heard to 12 practice sentences, and then to five blocks of 54 trials, wherein sentences from all three conditions plus the corresponding fillers were mixed in random order within each block. Each block lasted approximately 10 min. During the experiment, participants were instructed to listen to the sentences attentively for comprehension.

2.1.4. EEG recording

The EEG signal was recorded in a silent room during auditory stimulation from 31 passive channels mounted in an elastic cap. The channels were distributed over the head surface according to the 10% standard system of the American Electroencephalographic Society (see Fig. 1). Eye movements were monitored with two electrodes placed close to the right eye. The on-line reference electrode was attached to the tip of the nose. The activity over the right and left mastoids was also measured by two electrodes. Electrode impedance was kept below 5 kOhm. The EEG signal was filtered on-line with a 0.1–100 Hz bandpass filter and digitized at 500 Hz. EEG epochs started 100 ms before and lasted until 700 ms after the onset of the final word of the sentences. They were averaged for each participant, each experimental condition and each electrode. EEG epochs were accepted under an artefact rejection criterion of $-/+70$ μV at any channel (i.e. this criterion was applied by calculating the difference amplitude from one time frame to the next within the EEG epoch). The number of accepted EEG epochs did not differ across experimental conditions (fully expected, 42.7; phonologically unexpected, 42.3; fully unexpected, 42.6; $F(2, 40) = 0.45$, $p > .2$). The EEG signal was filtered offline by a

1 – 30 Hz bandpass filter and a 50 Hz notch filter. A 100-ms pre-stimulus baseline was also applied. For each participant, bad channels were interpolated (Perrin, Pernier, Bertrand, Giard, & Echallier, 1987) and the initial reference was changed offline to the average mastoid reference (left and right).

2.1.5. ERP analyses

The analyses[3] focused on the two time windows coinciding in latency with two negative ERP components associated with spoken word recognition, the N200 and N400. Based on the known properties of these components and on visual inspection of the waveforms, the two time windows were selected to explore the mean amplitude of each putative component across experimental conditions as follows: early time window (285–335 ms), late time window (350–600 ms). The latency window for the early time window corresponds to the onset of the ascending flank and the offset of the descending flank observed at Cz, where the maximum peak amplitude is found (putatively, the N200). For the late time window, the window contains the maximum peak amplitude at Pz and large standard time range to study N400 effects (Van den Brink & Hagoort, 2004; van den Brink et al., 2001). A three-way repeated measures ANOVA was performed on the mean ERP amplitude in each critical time window with factors: Condition (3: Fully expected, Phonologically unexpected, Fully unexpected), Anterior/Posterior (2: Anterior, Posterior) and Laterality (3: Left, Midline, Right). Eighteen scalp sites were included for the statistical analysis, grouping by Anterior/Posterior position and Laterality as follows: Left Anterior: F3, FC5, C3; Midline Anterior: Fz, FC1, FC2; Right Anterior: F4, FC6, C4; Left Posterior: P3, CP5, PO1; Midline Posterior: CP1, CP2, Pz; Right Posterior: P4, CP6, PO2. Scalp sites and the two factors Anterior/Posterior and Laterality were chosen to provide appropriate scalp coverage to study the components of interest (Kutas & Federmeier, 2000; Van den Brink & Hagoort, 2004). When there was more than one degree of freedom in the numerator, the Greenhouse-Geisser correction was applied to adjust for violations of sphericity (Greenhouse & Geisser, 1959); the corrected p values are reported.

2.2. Results

Grand-average waveforms[4] in each experimental condition and topographical maps of effects are displayed in Fig. 2. In the early time window, there was a significant main effect of condition [$F(2, 40) = 9.98$, $p < .001$] and a significant Condition × Anterior/Posterior interaction [$F(2, 40) = 3.46$, $p = .05$]. As seen in Fig. 2, the amplitude of the negativity over this time window was greater in the fully unexpected condition than both the fully expected and the phonologically unexpected conditions, particularly at posterior sites (fully unexpected vs. expected, [$F(1, 20) = 23.76$, $p < .001$]; fully unexpected vs. phonologically unexpected, [$F(1, 20) = 7.52$, $p < .05$]). The increased negativity in the fully unexpected condition with respect to the fully expected condition was stronger at the posterior

sites than at the anterior sites [$F(1, 20) = 4.52$, $p < .05$], and relative to the phonologically unexpected condition it did not reach significance at anterior sites [$F(1, 20) = 3.19$, $p = .09$]. Interestingly, the phonologically unexpected condition elicited a larger amplitude negativity with respect to the fully expected condition at posterior sites [$F(1, 20) = 7.13$, $p < .05$], whereas no significant effect was found at anterior sites [$F(1, 20) = 3.91$, $p = .07$]. The scalp distribution of the condition effect also depended on laterality, as shown by the significant Condition × Laterality interaction [$F(4, 80) = 3.70$, $p < .05$]. In particular, we found an increase in the amplitude of the negativity elicited by the phonologically unexpected condition relative to the fully expected condition at left and midline sites (left, [$F(1, 20) = 9.3$, $p < .01$], midline, [$F(1, 20) = 7.26$, $p < .05$]), but not over the right hemiscalp [$F(1, 20) = 2.81$, $p = .11$] (see Fig. 2).

Over the late time window, similarly to the early one, a main effect of condition was observed [$F(2, 40) = 20.23$, $p < .001$]. However, this time, only the fully unexpected condition elicited a significantly larger negativity than both the fully expected and the phonologically unexpected conditions (fully unexpected vs. expected, [$F(1, 20) = 22.67$, $p < .001$]; fully unexpected vs. phonologically unexpected, [$F(1, 20) = 37.19$, $p < .001$]). No significant difference in amplitude of the negativity appeared between the fully expected and the phonologically unexpected conditions [$F(1, 20) = 0.23$, $p > .2$]. The analyses also revealed a significant Condition × Laterality interaction [$F(4, 80) = 2.87$, $p = .05$]. This was explained by different scalp distribution elicited by the experimental conditions. While the fully unexpected condition yielded a negative wave centered on the midline sites (left vs. midline, [$F(1, 20) = 6.46$, $p < .05$], right vs. midline, [$F(1, 20) = 10.31$, $p < .01$]), the voltage values did not differ as a function of the laterality in the two other conditions. This topographical pattern for the fully unexpected condition was observed only at posterior sites (left vs. midline, [$F(1, 20) = 15.63$, $p < .01$], right vs. midline, [$F(1, 20) = 20.35$, $p < .01$]) as suggested by the significant Condition × Laterality × Site interaction [$F(4, 80) = 3.85$, $p < .05$].

In sum, the fully unexpected condition produced an early negative effect followed by a late effect with respect to the fully expected condition, whereas the phonologically unexpected condition produced an early negative effect but no late effect (see Table 2). These results of early and late negative shifts, corresponding to phonological and semantic mismatch respectively, are clearly reminiscent of the N200 and N400 components discussed in the introduction. Yet, as also pointed out earlier, the dissociation of N200 and N400 components has been somewhat controversial in the literature. In our data from the semantically anomalous condition, the late negative difference, corresponding to the typical N400 and reflecting semantic anomaly, had a slightly more posterior spatial distribution than the early negative difference (see Fig. 2). This topographical difference between the early and late effects is in line with the spatial distribution of the N200 and N400 waves described earlier by van den Brink and Hagoort (2004) as well as other studies (Connolly & Phillips, 1994; Van den Brink et al., 2001). However, these apparent topographical differences in our data were not sufficiently confirmed by statistical tests when comparing the mean amplitudes of effects between the early and late latency windows elicited by the semantically anomalous condition from all electrodes across the scalp after normalization to the global field power ($F(1, 20 = 0.25$, $p > .2$). Although this is tangential to our main point of interest, we note that Experiment 1 does not provide clear-cut evidence for two separable components. In consequence, we will refer hereafter to early and late effects associated with the phonological or semantic mismatch under the general label of N400 rather than in terms of two distinct ERP components.

Nevertheless, one potential concern with the interpretation of the present data is that critical differences between fully and

---

[3] In Experiment 1, wherein the sentential context was produced in the native accent of participants, a later latency and weaker amplitude of the N100 wave was obtained for the non-native phonological word forms relative to the native phonological word forms. Note that this same pattern was found for the non-native phonological word forms during the first half of Experiment 2. Here, the N100 triggered by the first phoneme and the cues related to the second phoneme was modulated by the familiarity of initial phonemes to produce the activation of lexical candidates based on the speech signal of the incoming word.

[4] The Cz electrode is shown in Figs. 2 and 3, as it was the electrode for which the amplitude of the early negativity was the greatest. Statistical analyses on the mean amplitude of Cz revealed a modulation of the amplitude of the early negativity elicited by the phonologically and fully unexpected conditions with respect to the fully expected condition, and a modulation of the amplitude of the late negativity elicited by the fully unexpected condition in Experiment 1. Contrary to this, in Experiment 2 only the fully unexpected condition elicited a modulation in the amplitude of the two negative responses.
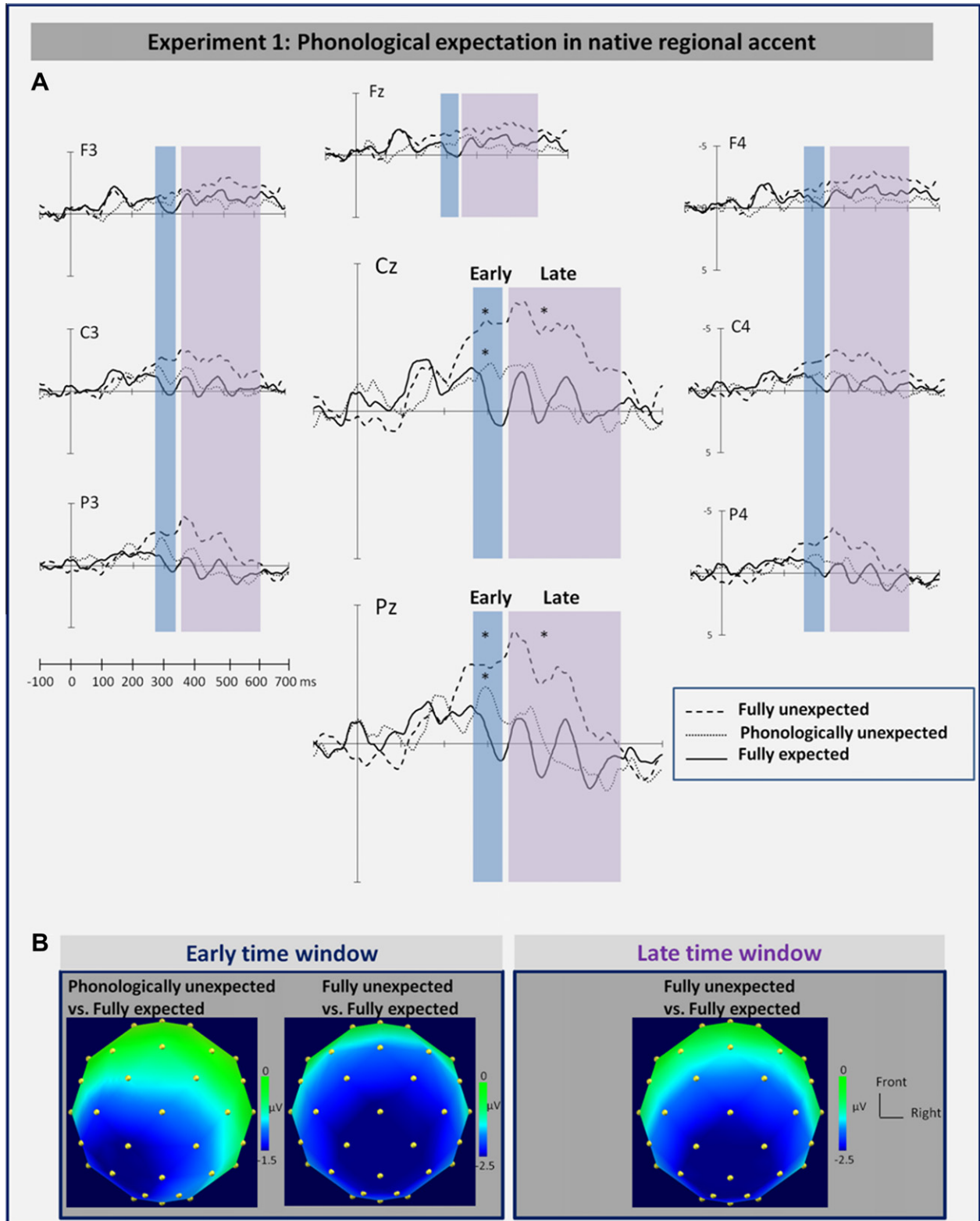
Fig. 2. (A) Grand-average waveforms for three conditions (fully expected, phonologically unexpected, Fully unexpected) in native regional context (Experiment 1). (B) Subtraction maps illustrating the differences between the fully unexpected condition or the phonologically unexpected condition and the fully expected condition. Starts (∗) denote significant effects.

phonologically unexpected conditions with respect to the fully expected condition in Experiment 1 could be contaminated by the differential presence of oscillatory activity in the alpha band (~10 Hz). We therefore decided to re-run the EEG data analyses under two variations; (1) removing *alpha* oscillatory

activity with an off-line filter (8–12 Hz) and (2) using a longer time window (100 ms) where the phase of the alpha oscillations is less likely to affect the results of tests for the early effects. In both cases, the pattern of results was confirmed (see Table 3).

Table 2
Summary of the observed ERP effects.

		Phonologically unexpected vs. fully expected	Fully unexpected vs. fully expected
EXPERIMENT 1	Early effect	+	+
	Late effect	−	+
EXPERIMENT 2	Early effect	−	+
	Late effect	−	+

+: Significant effects; −: no effects.

In conclusion, in Experiment 1 the early effect elicited by the phonologically unexpected condition suggests that phonological detail is taken into account in the mechanism generating phonological predictions about expected words.

3. Experiment 2: Phonological expectations in non-native regional accent

To provide more evidence regarding the context-based generation of expectancies for detailed word forms, we investigated whether the prediction of phonological forms arises when the contextual frame is presented in the non-native regional accent of the participant. The question here is whether listeners will adapt their expectancies to the less familiar non-native pronunciation of the context. In this case, if predictive mechanisms can flexibly adapt according to prior phonological context, then it is the native pronunciation of the target that should result in a breach of phonological expectation.

3.1. Methods

Twenty-one newly selected students with the same features as those tested in Experiment 1 participated in Experiment 2. Note that all these participants (like most Eastern Catalan speakers) had knowledge of the regional variation in vowel realization in Western Catalan. The experimental stimuli used in Experiment 2 were the same as in Experiment 1, except that they were recorded again with a native Western Catalan speaker. Now, the native accent pronunciation of the target word corresponded to the phonologically unexpected condition, whereas the fully expected condition contained the non-native version of the target word. As in Experiment 1, the fully unexpected condition contained an accent-neutral, but semantically anomalous, target word. The recording and selection of these new sentences were carefully performed to make sure that intonation and speaking rate were kept constant within experimental conditions. A native Catalan-dominant female from Barcelona checked the number of phonemes marking the non-native regional accent in the context frames (mean: 6.5, range: 2–15) to ensure that the context clearly revealed the Western Catalan accent of the speaker. Additionally, we checked that the pronunciations of final words were correct according to the properties of experimental conditions. Similarly to Experiment 1, the total duration of the context sentences before the final target word and that of the final word was similar across experimental conditions ($F(2, 268) = 0.23$, $p > .2$; $F(2, 268) = 1.72$, $p = 0.19$, respectively). The total mean duration of sentences before the final word was 3323 ms (SD: 877) for the fully expected condition, 3380 ms (SD: 859) for the phonologically unexpected condition and 3357 ms (SD: 873) for the fully unexpected condition. The mean duration of final word was 452 ms (SD: 83) for the fully expected condition, 461 ms (SD: 91) for the phonologically unexpected condition and 467 ms (SD: 80) for the fully unexpected condition. The experimental procedure and the EEG set-up in Experiment 2 were identical to Experiment 1. For the ERP analysis, we focused on early and late time windows. Based on the visual inspection of waveforms, the two time windows were selected to explore the mean amplitude over various timing across experimental conditions as follows: 285–335 ms and 350–600 ms. As in Experiment 1, a three-way repeated measures ANOVA was performed in each critical time window with factors: Condition (3: Fully expected, Phonologically unexpected, Fully unexpected), Anterior/Posterior (2: Anterior, Posterior) and Laterality (3: Left, Midline, Right).

Table 3
Statistical results for the new EEG data analyses in Experiment 1.

Early time window Significant main effects or interactions	Filtered data	Significant main effects or interactions	Unfiltered data (100 ms)
Condition effect	$F(2, 40) = 13.38$, $p < .001$	Condition effect	$F(2, 40) = 15.34$, $p < .001$
Condition × Anterior/Posterior interaction	$F(2, 40) = 9.75$, $p < .001$	Condition × Anterior/Posterior interaction	$F(2, 40) = 8.87$, $p < .001$
Over Posterior sites		Over posterior sites	
Fully unexpected vs. expected	$F(1, 20) = 28.65$, $p < .001$	Fully unexpected vs. expected	$F(1, 20) = 32.18$, $p < .001$
Fully unexpected vs. Phonologically unexpected	$F(1, 20) = 16.82$, $p < .001$	Fully unexpected vs. Phonologically unexpected	$F(1, 20) = 23.29$, $p < .001$
Phonologically unexpected vs. fully expected	$F(1, 20) = 5.27$, $p < .05$	Phonologically unexpected vs. Fully expected	$F(1, 20) = 5.31$, $p < .05$
Condition × Laterality interaction	$F(4, 80) = 4.09$, $p < .05$	Condition × Laterality interaction	$F(4, 80) = 3.75$, $p < .05$
Phonologically unexpected vs. expected		Phonologically unexpected vs. expected	
At left sites	$F(1, 20) = 9.14$, $p < .01$	At left sites	$F(1, 20) = 7.59$, $p < .05$
At midline sites	$F(1, 20) = 7.07$, $p < .05$	At midline sites	$F(1, 20) = 6.50$, $p < .05$
At right sites	$F(1, 20) = 2.21$, $p = .15$	At right sites	$F(1, 20) = 2.45$, $p = .13$
Late time window – filtered data			
Condition effect	$F(2, 40) = 20.88$, $p < .001$		
Fully unexpected vs. expected	$F(1, 20) = 25.19$, $p < .001$		
Fully unexpected vs. Phonologically unexpected	$F(1, 20) = 36.91$, $p < .001$		
Phonologically unexpected vs. Fully expected	$F(1, 20) = 0.11$, $p > .2$		
Condition × Laterality interaction	$F(4, 80) = 2.60$, $p < .05$		
Fully unexpected, Left vs. Midline	$F(1, 20) = 6.92$, $p < .05$		
Fully unexpected, Right vs. Midline	$F(1, 20) = 9.33$, $p < .01$		
Condition × Laterality × Site	$F(4, 80) = 3.88$, $p < 0.05$		
Over Posterior sites			
Fully unexpected, Left vs. Midline	$F(1, 20) = 16.06$, $p < .001$		
Fully unexpected, Right vs. Midline	$F(1, 20) = 18.71$, $p < .001$		

3.2. Results

Grand-average waveforms in each experimental condition and topographical maps of effects are displayed in Fig. 3. The amplitude of the early negativity was modulated by the experimental conditions as indicated by the main effect of condition [$F(2, 40) = 23.56$, $p < .001$]. As seen in Fig. 3, the fully unexpected condition led to a greater amplitude of the early negativity in comparison to both the fully expected and phonologically unexpected conditions (fully unexpected vs. fully expected, [$F(1, 20) = 29.55$, $p < .001$]; fully unexpected vs. phonologically unexpected, [$F(1, 20) = 36.27$, $p < .001$]). However, in this experiment, there was no significant difference in the amplitude of the early negativity between the fully expected and phonologically unexpected conditions [$F(1, 20) = 0.07$, $p > .2$]. The increase in the amplitude of the early negativity elicited by the fully unexpected condition presented differential scalp distribution as suggested by the significant Condition × Anterior/Posterior interaction [$F(2, 40) = 15.54$, $p < .001$]. In particular, this increase in early negativity amplitude was larger at posterior sites relative to anterior sites (fully unexpected vs. fully expected, [$F(1, 20) = 21.25$, $p < .001$]; fully unexpected vs. phonologically unexpected, [$F(1, 20) = 22.48$, $p < .001$]). There was also a significant Condition × Laterality interaction [$F(4, 80) = 2.99$, $p < .05$]. The increase in early negativity amplitude elicited by the fully unexpected condition was greater over midline sites than over left and right sites (fully unexpected vs. fully expected and midline vs. left, [$F(1, 20) = 23.83$, $p < .001$]; fully unexpected vs. fully expected and midline vs. right, [$F(1, 20) = 7.8$, $p < .05$]; fully unexpected vs. phonologically unexpected and midline vs. left, [$F(1, 20) = 6.5$, $p < .05$]; fully unexpected vs. phonologically unexpected and midline vs. left, [$F(1, 20) = 10.04$, $p < .01$]).

The analyses of the large late negativity revealed a main effect of condition [$F(2, 40) = 35.64$, $p < .001$], showing that the fully unexpected condition yielded a greater amplitude of the late negativity in comparison to both the fully expected and phonologically unexpected conditions (fully unexpected vs. fully expected, [$F(1, 20) = 38.33$, $p < .001$]); fully unexpected vs. phonologically unexpected, [$F(1, 20) = 61.35$, $p < .001$]). Like the early negativity, the modulation in the amplitude of the late negativity showed a specific scalp distribution, with stronger amplitudes at posterior sites relative to anterior sites (fully unexpected vs. fully expected, [$F(1, 20) = 27.33$, $p < .001$]; fully unexpected vs. phonologically unexpected, [$F(1, 20) = 24.59$, $p < .001$]). In addition to a significant Condition × Anterior/Posterior interaction [$F(2, 40) = 18.76$, $p < .001$], a significant Condition × Laterality interaction was found [$F(4, 80) = 5.25$, $p < .01$]. The increase in late negativity amplitude elicited by the fully unexpected condition was greater over midline sites than over left and right sites (fully unexpected vs. fully expected and midline vs. left, [$F(1, 20) = 13.14$, $p < .01$]; fully unexpected vs. fully expected and midline vs. right, [$F(1, 20) = 11.95$, $p < .01$]; fully unexpected vs. phonologically unexpected and midline vs. left, [$F(1, 20) = 9.66$, $p < .01$]; fully unexpected vs. phonologically unexpected and midline vs. left, [$F(1, 20) = 7.11$, $p < .05$]). This latter effect also depended on the Anterior/Posterior factor as indicated by the significant Condition × Laterality × Anterior/Posterior interaction [$F(4, 80) = 7.14$, $p < .001$]. While the increase of late negativity amplitude elicited by the fully unexpected condition was greater over midline sites than over left and right hemiscalp at posterior sites, this increase of late negativity amplitude was greater over midline and right hemiscalp at anterior sites (posterior: fully unexpected vs. fully expected and midline vs. left, [$F(1, 20) = 12.99$, $p < .01$]; fully unexpected vs. fully expected and midline vs. right, [$F(1, 20) = 27.79$, $p < .001$]; fully unexpected vs. phonologically unexpected and midline vs. left, [$F(1, 20) = 9.55$, $p < .01$]; fully unexpected vs. phonologically unexpected and midline vs. left, [$F(1, 20) = 20.57$, $p < .001$]; anterior: fully unexpected

vs. fully expected and midline vs. left, [$F(1, 20) = 11.40$, $p < .01$]; fully unexpected vs. fully expected and right vs. left, [$F(1, 20) = 4.24$, $p = .05$]; fully unexpected vs. phonologically unexpected and midline vs. left, [$F(1, 20) = 7.86$, $p < .05$]; fully unexpected vs. phonologically unexpected and right vs. left, [$F(1, 20) = 4.05$, $p = .05$]). Note that similarly to Experiment 1, no topographical difference between the early and late effects elicited by the semantically anomalous condition was found ($F(1, 20) = 0.14$, $p > .2$). The mean ERP amplitude in early and late latency windows for each condition over posterior sites where the effects were the highest is displayed in Fig. 4 (at the top, Experiment 1, at the bottom, Experiment 2).

To summarize, in Experiment 2, wherein the contextual frame was presented in a non-native regional accent, the fully unexpected condition produced an early and late negative deflection sequence with respect to the fully expected condition, just as in Experiment 1. However, in this case the phonologically unexpected condition did not produce an early negativity effect (nor a late negativity effect) and its ERP response was thus equivalent to that of the fully expected condition (see Fig. 3 and Table 2). This means that the regional variation was in this case not treated as a mismatch, and, therefore, that the expected word form was not specified in detail.

4. Control experiment: Categorization of the critical vowels of target words

Contrary to what was observed in Experiment 1, the phonologically unexpected condition elicited no modulation in the early negativity effect in Experiment 2. This was reinforced by a significant interaction (unfiltered data, $F(1, 40) = 4.99$, $p < .05$; filtered data, $F(1, 40) = 4.49$, $p < .05$) between the mean amplitude over the early negativity in the fully expected and phonologically unexpected conditions across the two experiments. However, one trivial possible explanation for this lack of effect could be that categorization of the critical phoneme in the final target word was easier in Experiment 1 than in Experiment 2, leading to more efficient detection of the phonological mismatch (expected vs. unexpected word form) in the former experiment. More specifically, the difference in speaker used for the materials of Experiments 1 and 2 (required for the regional accent change) could have affected fine-grained acoustic cues, and thereby the participants' ability to categorize the critical phonemic variation. We therefore examined, in a separate behavioral experiment, the listeners' ability to categorize these critical phonemes in an isolated presentation of the target words. All target words (always in sentence final position) were extracted from the soundtrack of the sentences and presented to native Catalan-dominant students from the University of Pompeu Fabra, with the same demographic profile as those tested in Experiments 1 and 2. Different participants were tested with each set of materials (Eastern Catalan speaker and Western Catalan speaker; $n = 15$ each). All participants were asked to attentively listen to the words and categorize the first vowel of the target words by button press in a speeded 2AFC task. We measured accuracy and response time (see Table 4) according to the regional accent of the target word (native and non-native word forms), and the speaker pronouncing the words (Eastern Catalan speaker or Western Catalan speaker, in Experiments 1 and 2, respectively). Participants responded rather accurately and, importantly, performance did not vary as a function of the speaker [$F(1, 28) = 0.54$, $p > .2$] or regional accent pronunciation [$F(1, 28) = 0.15$, $p > .2$]. There was no significant interaction between the speaker and regional accent pronunciation factors [$F(1, 28) = 0.0001$, $p > .2$]. In terms of response times, participants responded more quickly to words with a non-native pronunciation compared to those with a native one
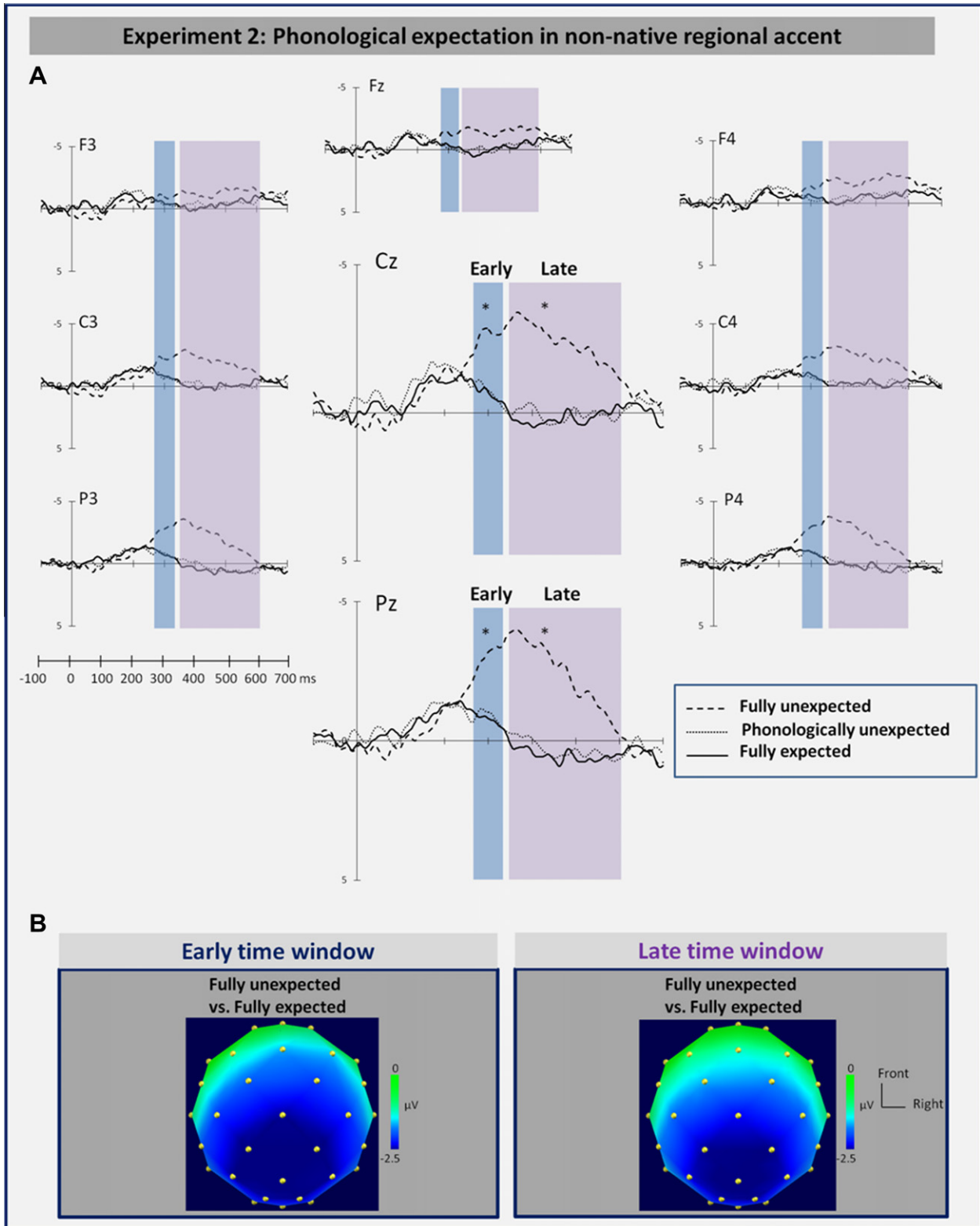
Fig. 3. (A) Grand-average waveforms for three conditions (fully expected, phonologically unexpected, fully unexpected) in non-native regional context (Experiment 2). (B) Subtraction maps illustrating the differences between the fully unexpected condition and the fully expected condition. Stars (∗) denote significant effects.

$[F(1, 28) = 11.67, p < .01]$, but critically this difference was equivalent for the two sets of materials (i.e., those used in the two experiments) and therefore, was not particularly influenced by the different speakers used in the ERP experiments $[F(1, 28) = 2.91, p = .12]$.

In light of the results of this control experiment, then, it is unlikely that the differences in ERP pattern across Experiments 1 and 2 are explained by differences in categorization or saliency of the critical vowel in the final target words.
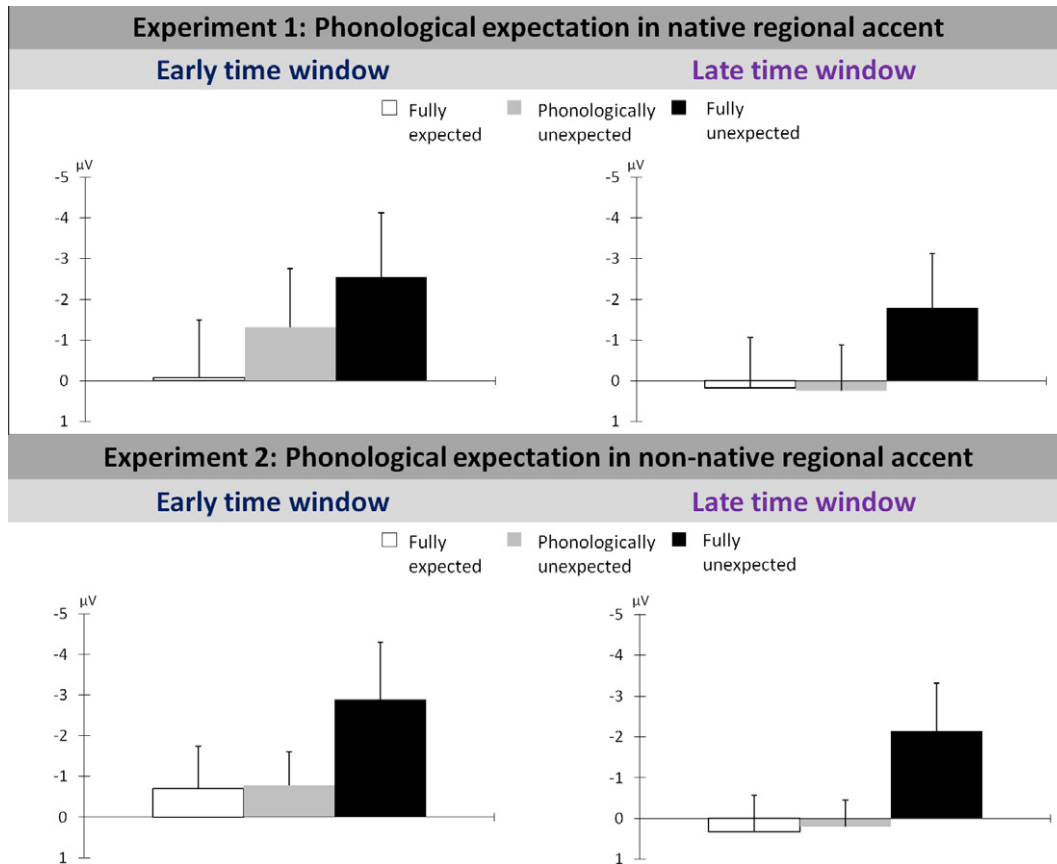
Fig. 4. Mean ERP amplitude in early and late latency windows for each condition (fully unexpected, phonologically unexpected and fully expected) in Experiments 1 and 2 over posterior sites (where the effects were strongest).

Table 4
Mean accuracy (in%) and reaction times (RT in ms) in control experiment.

	BCN speaker: stimuli of Experiment 1		Lleida speaker: stimuli of Experiment 2	
	Native word forms	Non-native word forms	Native word forms	Non-native word forms
Accuracy	93.2 (7.9)	92.7 (5.4)	91.7 (7.1)	91.2 (5.1)
RTs	1018 (149)	959 (138)	1062 (181)	1042 (183)

Standard deviations in parentheses, BCN speaker: Barcelona speaker, i.e., Eastern Catalan accent, Lleida speaker: Western Catalan speaker.

5. Discussion

The present study investigated whether word-form predictions are specified in accordance with the phonological variability of the context during online spoken word comprehension. We measured ERPs evoked by contextually expected words pronounced in either an expected or an unexpected regional accent. To do so, the native vs. non-native accent of the carrier context sentence was manipulated in two experiments, so that we could decouple the effects of phonological expectancy from the phonological familiarity (native vs. non-native) of the target word. The experimental logic was that, if the listeners expect specific regional phonological word forms, then the target word in the regional accent that does not match that of the carrier sentence should have an impact at early processing stages reflecting the phonological mismatch compared to the completely expected word form. The study showed that target words pronounced in an unexpected regional accent elicited a larger early negative response (~250 ms) than the expected regional word form when the carrier context sentence was spoken in native accent, whereas no modulation of this early negativity appeared

for the same comparison when the context sentence was in non-native accent. Interestingly, the pattern of results was thus consistent with the specific phonological prediction hypothesis when the sentence context that listeners were exposed was spoken with their native accent, but with the phonologically under-specified prediction hypothesis when they were presented with sentence frames in a non-native accent. Whatever the regional accent of the context, and as expected from previous studies, both the early and late negative effects were clearly seen in the fully unexpected condition, signalling the sensitivity of our paradigm to phonological and semantic mismatch, with respect to the two different realizations of the semantically acceptable word ending. The implications of these findings merit attention, and we discuss them below.

The result concerning the fully unexpected word from the meaningful preceding context is substantially comparable to the findings of previous studies introducing phonological and semantic mismatch in listeners' native accent (Connolly & Phillips, 1994; Van den Brink & Hagoort, 2004; Van den Brink et al., 2001). Indeed, we replicate the original findings of Connolly and Phillips (1994)

showing that the eliciting of a sustained amplitude of negative responses at early and late time windows to words that were semantically anomalous and whose initial phonemes were unexpected given the preceding context. Similarly to van den Brink and Hagoort (2004), the scalp distribution of the early negativity effect triggered by the semantically anomalous condition did not statistically differ from that of the late negativity effect. These findings thus offer no conclusive evidence in favor of separating N200 effects from N400 effects. This result is particularly in line with the findings of Van Petten et al. (1999), which showed that the onset of N400 effects is related to the moment at which there is sufficient incoming signal to register a mismatch. This supports the interpretation of the N400 effect as a congruency effect associated with both phonological and semantic mismatches.

Interestingly, our study also showed that the regional accent in which the context was spoken did not modulate the generation of upcoming word expectations based on semantic constraints from the prior context. That is, the early and late pattern effects arose clearly in both experiments for the fully unexpected condition, and in consequence, it does not appear to depend much on the particular regional accent being spoken.[5] Although the context seems to generate a lexical expectation about the upcoming word independently of the regional accent presented in the context, the important question here concerns whether particular detailed phonological word forms are specified in accordance with the regional accent presented in the context.

According to the hypothesis laid out in the introduction, the early effects associated with the processing of critical vowels made it possible to probe whether particular detailed phonological word forms are expected as a function of the regional accent presented in the context. The results of each experiment were clear, but the pattern differed as a function of whether the sentence context was spoken in the native accent or in the non-native variety. In particular, when the contextual frame was in the native accent, the results clearly argue in favor of a model incorporating detailed phonological forms. However, when the sentential context was spoken in a non-native regional accent, no differences between expected and unexpected phonological word forms were found. This latter result might in principle be accounted for by the alternative framework, with listeners producing phonologically under-specified predictions, so that the phonological acceptability of the two phonological word forms is equivalent.

What does this pattern of results say about the contrasting models of word representation discussed in the introduction? First of all, and as mentioned earlier, the present results are relatively neutral with respect to the question of whether the level of specification of predictions comes down to the level of phonetic detail. This is because the regional variation exploited in these experiments relates to phonologically meaningful categories. That said, it is logical that the hypothesis of prediction of upcoming words in particular phonological forms is, at least, consistent with exemplar-based models of the lexicon (Goldinger, 1998; Johnson, 1997) suggesting the storage of pronunciation variants such as those that occur in different regional accents. Yet, the contrasting pattern of results between native and non-native context is somewhat unsupportive of a strong version of any of the two models, either exemplar-based or abstract representation. Therefore, an account that encompasses the present results as a whole must resort to a more flexible representational model. Along these lines, some recent models of spoken word representation, referred to as hybrid models, such as for example, the frequency-based model (Connine, Rambon, & Patterson, 2008), have been proposed. In particular,

Connine et al. (2008) proposed that main phonological variants of a word are jointly stored in the mental lexicon. On this account, the frequency of occurrence of each variant is encoded in memory (note that the encoding of the frequency of occurrence of words' sound shape was already advanced by Goldinger (1998). In line with the frequency-based model, therefore, the phonologically under-specified prediction occurring in an unfamiliar accent context could be explained by the low frequency and stability of phonologically expected but unfamiliar word forms in contrast to the high frequency of the phonologically unexpected but familiar word forms. Hence, listeners could have expected both the likely but unfamiliar word forms and the phonologically unexpected but familiar word forms to a similar extent. A plausible alternative hypothesis for the present results would be a flexible predictive mechanism that adapts the level of specification according to prior knowledge. In particular, when the context is familiar, the mechanism can draw upon detailed phonological knowledge to project highly specified phonological expectations. Such highly specified phonological expectations would provide fine sensitivity to mismatch between the expected and incoming forms. This is in line with the ERP-graded effect of similarity to the expected form observed in familiar context, given that the fully unexpected condition, which was acoustically more different from the fully expected condition, produced a larger ERP shift than the phonologically unexpected condition, obviously more similar from an acoustic point of view. Instead, according to this alternative hypothesis, when the context is phonologically unfamiliar, predictions cannot be fine-tuned to a detailed form because the priors for the prediction are less precise. In this case the system would default to a less specified prediction mode. Contrary to the mechanism proposed earlier, this flexible predictive mechanism would easily support the possibility that the expectancies might express at a phonetic level.

This kind of context-based flexibility is certainly possible if one conceptualizes predictive mechanisms in a framework where constraints occur within but also between different levels of speech representation (from semantics to phonology; see Pickering & Garrod, 2007). In line with this idea, Hanulíková, van Alphen, van Goch, and Weber (2012) proposed that listeners adjust their probability model to ensure successful communication in a non-native accent, such as a foreign accent. These authors demonstrated that knowledge of the speaker's characteristics can modulate the neural correlates of syntactic processing. When native listeners were exposed to gender violations in native speech input, a P600 effect was observed, but when they listened to the same violations produced by a non-native speaker, no ERP effect was found. Overall, it thus appeared that listeners adjust the probability model in taking into account speaker's characteristics during spoken language comprehension (see also, van Berkum et al., 2008).

Moreover, in addition to predictive mechanisms whose level of specification depends on the accent of the prior context, which we may call global expectancies, local expectancies based on the familiarity of word-form pronunciations could also occur independently of the context. Consequently, when the context is familiar, the early effect could result from the addition of the effects of mismatches based on both global and local expectancies. When the context is not familiar, however, the interaction of mismatches based on global and local expectancies would produce antagonistic effects. This could thus explain the lack of differences between expected and unexpected phonological word forms in the unfamiliar context. However, even if an impact of local expectancies cannot be totally excluded, the involvement of predictive mechanisms depending on the accent of the prior context is the sole mechanism necessary to explain the pattern of results.

Interestingly, in auditory cognitive neuroscience, it has been recently proposed that the cortical organization of hierarchical

---

[5] Additional analyses showed that the increase in the amplitude of early and late negative effects elicited by the fully unexpected condition was not influenced by the regional accent of the sentential context ($p > .2$).

sensory systems via predictive coding subtends perceptual learning and inference that is, that each level of the hierarchy receives incoming sensory input from the level below and top-down predictions from the level above (Baldeweg, 2006; Friston, 2005; Friston & Kiebel, 2009). Following this framework, our findings might provide evidence for brain mechanisms of predictive coding at the phonological/phonetic level shaped by regional accent. It seems that the activation of particular detailed word forms on the basis of the preceding context might depend on the regional accent of the context. Hence, the influence of contextual constraints on word recognition includes a prediction of word forms based on the phonological variability of the context. We have proposed two main possible mechanisms that could account for this pattern of findings: on one hand, hybrid representational models where variants of the lexical item are weighted by their frequency of occurrence; on the other hand, a flexible predictive coding model wherein word forms predictions are only as detailed as it is allowed by the degree of precision to which the context can be parsed.

Acknowledgments

References

Alarcos, E. (1953). Sistema fonemático del catalán. *Archivum*, III (pp. 135–146). *Reeditado en Estudis de lingüística catalana* (pp. 11–32). Barcelona: Ariel.

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*(3), 247–264.

Baldeweg, T. (2006). Repetition effects to sounds: Evidence for predictive coding in the auditory system. *Trends in Cognitive Sciences, 10*, 93–94.

Connine, C. M., Rambon, L. J., & Patterson, D. J. (2008). Processing variant forms in spoken word recognition: The role of variant frequency. *Perception and Psychophysics, 70*, 403–411.

Connolly, J. F., & Phillips, N. A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *Journal of Cognitive Neuroscience, 6*, 256–266.

Connolly, J. F., Phillips, N. A., Stewart, S. H., & Brake, W. G. (1992). Event-related potential sensitivity to acoustic and semantic properties of terminal words in sentences. *Brain and Language, 43*, 1–18.

Connolly, J. F., Stewart, S. H., & Phillips, N. A. (1990). The effects of processing requirements on neurophysiological responses to spoken sentences. *Brain and Language, 39*, 302–318.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8*, 1117–1121.

Frauenfelder, U. H., & Tyler, L. K. (1987). The process of spoken word recognition: An introduction. *Cognition, 25*(1–2), 1–20.

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B, 360*, 815–836.

Friston, K., & Kiebel, S. (2009). Cortical circuits for perceptual inference. *Neural Networks, 22*, 1093–1104.

Giles, H., Coupland, N., & Coupland, J. (1991). Accommodation theory: Communication, context and consequences. In H. Giles, J. Coupland, & N. Coupland (Eds.), *Contexts of accommodation: Development in applied sociolinguistics* (pp. 1–68). Cambridge University Press.

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 22*, 1166–1183.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*, 251–279.

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika, 24*, 95–111.

Hanulíková, A., van Alphen, P. M., van Goch, M. M., & Weber, A. (2012). When one person's mistake is another's standard usage. The effect of foreign accent on syntactic processing. *Journal of Cognitive Neuroscience, 24*, 878–887.

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 145–166). San Diego, CA: Academic Press.

Kamide, Y., Altmann, G. T., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language, 49*, 133–156.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 12*, 463–470.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*, 161–163.

Liberman, A. M., Cooper, F. S., Harris, K. S., & MacNeilage, P. F. (1962). A motor theory of speech perception. *Proceedings of the speech communication seminar* (Vol. 2). Stockholm: Royal Institute of Technology.

Marslen-Wilson, W. D. (1984). Function and process in spoken word recognition: A tutorial review. In H. Bouma & D. G. Bouwhis (Eds.), *Attention and performance X: Control of language processes* (pp. 125–150). Hillsdale, NJ: Erlbaum.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continous speech. *Cognitive Psychology, 10*, 29–63.

McClelland, J. L., & Elman, J. L. (1986). The Trace model of speech perception. *Cognitive Psychology, 18*, 1–86.

Morton, J. (1979). Word recognition. In J. Morton & J. C. Marshall (Eds.), *Structures and processes* (pp. 108–156). Cambridge, MA: MIT Press.

Natale, M. (1975). Convergence of mean vocal intensity in dyadic communication as a function of social desirability. *Journal of Personality and Social Psychology, 32*, 790–804.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology, 47*, 204–238.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*, 97–113.

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America, 119*, 2382–2393.

Perrin, F., Pernier, J., Bertrand, O., Giard, M.-H., & Echallier, J. F. (1987). Mapping of scalp potentials by surface spline interpolation. *Electroencephalography and Clinical Neurophysiology, 66*, 75–81.

Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences, 11*, 105–110.

Rafel i Fontanals, J. (1998). Diccionari de freqüències. 3, dades globals [Frequency dictionary. 3, global data]. Barcelona: Institut d'Estudis Catalans.

Stevens, K. N., & Halle, M. (1967). Remarks on analysis by synthesis and distinctive features. In W. Walthen-Dunn (Ed.), *Models for the perception of speech and visual form* (pp. 88–102). Cambridge, MA: MIT Press.

Tanenhaus, M. K., & Lucas, M. M. (1987). Context effects in lexical processing. *Cognition, 25*(1–2), 213–234.

van Berkum, J. J., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 443–467.

van Berkum, J. J., van den Brink, D., Tesink, C. M., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience, 20*, 580–591.

Van den Brink, D., Brown, C. M., & Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *Journal of Cognitive Neuroscience, 13*, 967–985.

Van den Brink, D., & Hagoort, P. (2004). The influence of semantic and syntactic context constraints on lexical selection and integration in spoken-word comprehension as revealed by ERPs. *Journal of Cognitive Neuroscience, 16*, 1068–1084.

Van Petten, C., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory and Cognition, 25*, 394–417.

Wicha, N. Y. Y., Bates, E. A., Moreno, E. M., & Kutas, M. (2003a). Potato not Pope: human brain potentials to gender expectation and agreement in Spanish spoken sentences. *Neuroscience Letters, 346*(3), 165–168.

Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2003b). Expecting gender: An event related brain potential study on the role of grammatical gender in comprehending a line drawing within a written sentence in Spanish. *Cortex, 39*, 483–508.

Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience, 16*, 1272–1288.